

Segmentation d'Images Couleur : application à la cytologie des séreuses en vue d'un tri informatique cellulaire.

Color Images Segmentation : Application to serous effusions cytology for Computer Cellular Sorting

O. Lezoray¹, A. Elmoataz^{1,2}, H. Cardot¹, G. Gougeon¹, M. Lecluse³, H. Elie³, M Revenu²

¹ LUSAC : Laboratoire Universitaire des Sciences Appliquées de Cherbourg

² GREYC : Groupe de recherche en Informatique, Instrumentation et Image de Caen

³ Service d'Anatomie et de Cytologie Pathologiques du Centre Hospitalier Pasteur de Cherbourg

Lezoray Olivier

Service d'Anatomie et de Cytologie Pathologiques du Centre Hospitalier Pasteur de Cherbourg

50102 Cherbourg

ou LUSAC, Site Universitaire, BP 78, 50130 Octeville

Tél : (2) 33 20 76 31

E-mail : Lezoray@calvin.info.unicaen.fr

ou olezoray.lusac@chbg.unicaen.fr

Résumé

Le *screening* (étape de lecture d'une lame) est une activité humaine manuelle en cytologie. Mais elle est également subjective. Un système informatique semi-automatique pourrait contribuer à la détection des erreurs de *screening* grâce à une meilleure fiabilité. Nous décrivons la structure du système de tri cellulaire que nous mettons en place pour les effusions des séreuses et nous détaillons plus particulièrement la première étape : la segmentation. Celle-ci est basée sur des opérateurs morphologiques utilisant l'information couleur dans différentes représentations, ceci afin d'isoler les cytoplasmes et les noyaux des cellules. Nous présentons les stratégies de segmentation cytoplasmique et nucléaire : la ligne de partage des eaux contrainte par marqueurs en utilisant la couleur. Les résultats obtenus ont été évalués visuellement par des experts pluridisciplinaires : les cellules sont toutes segmentées et leurs bordures cytoplasmiques et nucléaires sont bien déterminées.

Mots clés

Analyse d'image, Couleur, Segmentation, Cytologie, Ligne de Partage des eaux.

Abstract

Screening (reading step of a slide) is a manual human activity which involves its subjectivity. A semi-automated computer-based system could contribute to the detection of screening errors by the way of a greater reability. We describe the stucture of the computer cellular sorting for serous effusions we are developping. The aim of this paper is to present the first step of our system : Segmentation. It is based on mathematical morphology principles using color information in different color spaces. We want to find the cytoplasmic and the nuclear boundary of our cells. Segmentation strategies for cytoplasm and nuclear are described : watersheds using various color information. Final results have been visually evaluated by experts : all cells are segmented and their cytoplasmic and nuclear boundary are precisly defined.

Keywords

Image Analysis, Color, Segmentation, Cytology, WaterShed.

1 Introduction

On distingue deux types d'examen dans les laboratoires d'anatomie et de cytologie pathologiques : l'examen histologique et l'examen cytologique. L'histologie est l'observation de la coupe microscopique d'un tissu, la cytologie est l'examen d'un étalement de cellules provenant de prélèvements. Nous allons nous intéresser plus particulièrement à l'examen cytologique. Les cellules, fixées et colorées, présentes sur une lame sont analysées par un cyto-pathologiste. Cette étape d'analyse d'une lame s'appelle le *screening*. Elle consiste en une analyse visuelle au microscope de toutes les cellules présentes sur une lame cytologique. Le but de cette étape est la détection de cellules anormales ou suspectes afin d'établir un diagnostic fiable et valide. Ce *screening* manuel des prélèvements cytologiques est difficile, fatigant, consommateur de temps et est décrit comme un travail « intense, particulièrement complexe dont le résultat repose sur le point de vue de l'être humain » [6]. Exigeant la plus grande concentration, le *screening* implique la recherche au microscope d'un nombre relativement faible de cellules anormales parmi les centaines ou milliers qui figurent sur la lame. De plus, la plupart des lames à analyser étant négatives (ne contenant pas de cellules anormales), un état d'accoutumance se produit et des erreurs de *screening* peuvent survenir. Les principales causes sont l'omission par diminution de la vigilance en raison de la monotonie du travail et surtout le grand nombre de lames à analyser dans la journée. Les erreurs de ce type sont considérées comme étant indissociables du procédé de *screening* manuel et elles sont responsables d'un grand nombre de faux-négatifs. Certains travaux font état de pourcentages de faux-négatifs allant de 5% à 50% [6] et les attribuent pour une grande part à des erreurs de *screening*. Certaines mesures ont été envisagées pour effectuer une revue de 10% des lames données négatives [14] mais il n'est pas apparu que ce genre de contrôle des lames élimine des erreurs dans le *screening* d'une lame : d'autres mesures sont nécessaires. Une solution à ce problème des erreurs de *screening* est l'automatisation de tout ou d'une partie de ce processus. Un système informatique peut réellement apporter un plus au cyto-pathologiste dans le domaine de l'assurance qualité ainsi qu'une analyse objective et quantitative qui permettrait d'éliminer la plupart des effets pervers du *screening* manuel. Ainsi, dès 1950 des tentatives d'automatisation partielle ou totale du procédé de *screening* apparaissent [17]. Ces systèmes utilisent l'analyse d'images afin de repérer les cellules présentes sur une lame. L'impact de ces nouvelles technologies fut très fort et l'Académie Internationale de Cytologie émit rapidement les spécifications qu'un système d'analyse d'image se doit de respecter [5]. La majorité des travaux qui ont été menés jusqu'à nos jours portent sur l'examen des frottis cervicaux-utérins qui est l'examen le plus courant en cytologie. On note une évolution dans la méthodologie adoptée par ces systèmes. Dans les années 70, les travaux de recherche se tournent vers des systèmes entièrement automatiques sans intervention humaine [14]. Mais la séparation entre les spécimens normaux et anormaux n'étant pas suffisamment efficace et reproductible, les recherches se sont tournées vers des systèmes semi-

automatiques. Leur rôle est de détecter les cellules anormales sur une lame et de les montrer au cyto-pathologiste pour qu'il puisse établir son diagnostic. Le système seconde le pathologiste dans son travail sans le remplacer : il laisse le jugement au spécialiste.

Plusieurs systèmes opérationnels ont déjà vu le jour (PAPNET [15], ROCHE [7] et NeoPath [11]). Ils opèrent tous sur des lames de frottis cervicaux. Deux d'entre eux (PAPNET et NeoPath) ont reçu l'accréditation de la Food and Drug Administration (FDA) leur permettant une utilisation clinique. Dans cette optique d'aide au diagnostic, nous nous proposons de réaliser un système d'aide au diagnostic utilisant le tri cellulaire par analyse d'image : A.R.C.T.I.C (Aide à la Recherche en Cytologie par le Tri Informatique Cellulaire). Mais, contrairement à ceux existants, nous allons opérer sur des prélèvements de liquides d'épanchement des séreuses (plèvre, péritoine, ...).

2 Notre Système : ARCTIC

Le système que nous développons est donc une forme d'aide au diagnostic et il demande une nouvelle forme d'examen de la lame de la part du cyto-pathologiste (Figure 1). Les lames qui seront données comme négatives par le cyto-technicien après son *screening* seront revues par notre système ARCTIC. Celui-ci fera un tri des cellules afin de vérifier que la lame était effectivement négative (ne contenant pas d'éléments suspects). Dans le cas inverse, les cellules suspectes omises seront affichées à l'écran pour une relecture par un pathologiste afin d'établir un diagnostic. Le système aidera donc le pathologiste en vue d'un diagnostic en recherchant les éléments rares suspects ou anormaux qui auraient pu échapper à la vigilance du cyto-technicien lors de son *screening* ; instaurant ainsi une baisse des faux-négatifs par la mise en place d'un contrôle qualité.

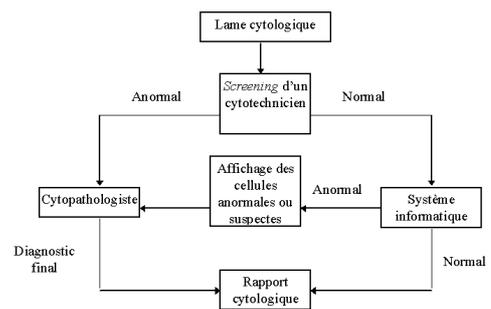


Figure 1 : l'examen d'une lame.

La mise au point d'un tel système requiert des connaissances d'horizons divers (analyse d'images, anatomopathologie, intelligence artificielle).

Notre système procède à l'analyse d'une image en trois étapes :

- 1 Segmentation : l'extraction des régions et des objets se trouvant sur l'image, à savoir les composantes cellulaires,
- 2 Caractérisation : l'évaluation des caractères extraits à la segmentation. Cela consiste à fournir des données représentatives pour chaque cellule,

③ Classification : l'étiquetage qui recherche la meilleure classification des régions obtenues lors de la segmentation. On se sert des caractéristiques cellulaires trouvées lors de l'étape précédente. Cela consiste en fait à trouver le type de la cellule et à affirmer sa bénignité ou sa malignité.

Des travaux utilisant cette architecture ont déjà été conçus par le GREYC avec lequel nous travaillons. Nous utilisons ces acquis pour mettre en place les trois étapes de notre système [13]. Nous venons actuellement de terminer la mise au point de la première étape de notre système et nous allons l'exposer dans la suite de cet article.

3 Matériels

Les étalements sont réalisés à partir de prélèvement de liquides d'épanchement des séreuses. Celui-ci est centrifugé et le culot de centrifugation est étalé sur une lame et fixé par un fixateur standard : l'alcool-acétone. Les cellules sont ensuite colorées par la méthode de Papanicolaou [10] (standard international de coloration). Elles ont alors un noyau coloré en bleu-violet par l'hématoxyline de Harris et un cytoplasme en vert ou bleu-vert par EA50. Les lames sont observées sur un microscope optique (OLYMPUS BX50) et sont enregistrées sur trois canaux RVB par une caméra 3-CCD Sony XC-003P connectée à une carte d'acquisition µTech implantée dans un PC Compaq Deskpro 6000 sous Windows NT. Les images ont une taille de 512x512 avec une dynamique de 256 niveaux de gris par composante. Nos images sont obtenues avec des conditions d'acquisition toujours constantes : grossissement de 20, diaphragme ouvert à 0,3, tension constante maintenue à 9V aux bornes du microscope. Afin de minimiser le risque de fluctuations dans l'alimentation en courant de notre système, tous les appareils sont alimentés par un courant ondulé. Nous développons ARCTIC avec Visual C++ 4.0 et la librairie de traitement d'images PANDORE [2] à laquelle nous avons rajouté un module permettant d'utiliser et de traiter des images couleurs.

4 Calibrage du système

Pour vérifier la linéarité dans le temps des images obtenues, nous avons réalisé la capture d'une même image (représentant un fond blanc uniforme) toutes les cinq minutes après avoir allumé simultanément microscope et caméra. Nous avons ensuite calculé la différence entre deux images successives, cette différence représente ce que nous appellerons le « bruit d'acquisition » d'une image à une autre. Nous avons alors pu quantifier ce bruit grâce aux images de différences et nous avons calculé la moyenne et la variance du bruit d'acquisition (en niveaux de gris) sur chacun des canaux R,V,B. Nous avons constaté que le bruit diminuait avec le temps. Si l'on modélise l'intensité associée à un pixel, on obtient la formule :

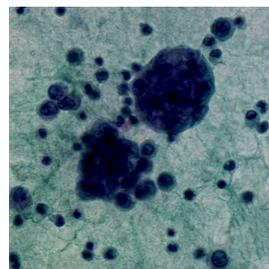
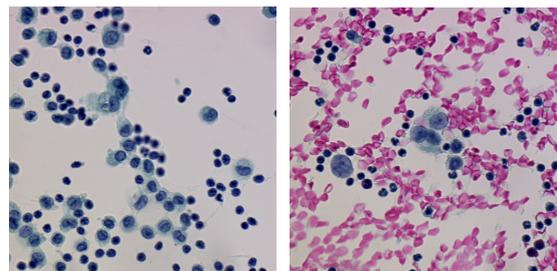
$$I_{lue} = I_{exacte} + I_{bruit} + I_{thermique}$$

I_{lue} est l'intensité lumineuse d'un pixel, I_{exacte} est l'intensité lumineuse qui devrait être donnée par le système si celui-ci était parfait, I_{bruit} est un biais additif constant qui représente le bruit généré par notre système, $I_{thermique}$ est le bruit

thermique additif provoqué par l'ensemble caméra et microscope. Au bout d'environ une heure, $I_{thermique}$ devient négligeable par rapport à I_{bruit} car on observe une stabilité du bruit qui n'est plus que le résultat de l'action de I_{bruit} sur I_{exacte} . Lorsque l'on fera une acquisition d'image, cela devra se faire au moins une heure après avoir allumé simultanément caméra et microscope, ceci afin d'atteindre leurs stabilités thermiques respectives. Une fois cette stabilité atteinte, nous avons pu caractériser le bruit I_{bruit} : il est représenté par une variable aléatoire gaussienne centrée. Ce bruit peut avoir une incidence de variation de 16 niveaux de gris par pixel soit environ 6,27%.

5 Description des images

Nous voulons isoler les cellules dans une image, c'est-à-dire reconnaître leurs cytoplasmes et noyaux : le cytoplasme pour avoir tout d'abord l'information de contexte (car nos cellules peuvent être aussi bien isolées qu'en amas) et le noyau pour apprécier la malignité d'une cellule. Les deux informations devront permettre de différencier tous les types cellulaires. Avant de segmenter nos images, il nous faut étudier les types de fond et d'objet que nous pouvons rencontrer.



Figures 2,3 et 4 : la variabilité des fonds, de la population et de l'organisation cellulaire.

On distingue les fonds clairs homogènes (figure 2), les fonds hémorragiques (figure 3 : globules rouges colorés en rouge) et les fonds muco-protéiques (figure 4 : mucus et protéines colorés en vert).

Les objets que nous pouvons rencontrer sont très variés : des cellules isolées (figure 2), accolées (figure 3) voire même en amas ou en papilles (figure 4). Notre but est d'isoler les noyaux et les cytoplasmes de toutes les cellules présentes sur l'image puis de séparer les cellules aux noyaux accolés (les cellules ayant des noyaux trop agglomérés ou se chevauchant sont considérées comme non séparables). Certaines cellules comme les hématies (globules rouges) ne présentent aucun intérêt au niveau du diagnostic, il n'est donc pas utile de les inclure dans notre stratégie de segmentation. Il nous faudra mettre en place une technique afin de les éliminer. La couleur apparaît alors comme un atout déterminant car d'une part on ne retrouve nulle part

ailleurs et pour aucune cellule la couleur rouge sur nos images et d'autre part les cytoplasmes sont verts et les noyaux bleus. Ainsi, une segmentation tirant parti des informations couleur intrinsèquement contenues dans les images nous paraît tout à fait appropriée. Une segmentation basée sur la couleur utilise les informations de ses trois composantes et ceci dans n'importe quel espace couleur : RVB, XYZ, HSL, $L^*u^*v^*$, $L^*a^*b^*$... [8,19]. Nous allons voir la méthodes de segmentation couleur que nous avons mise en œuvre pour isoler les cytoplasmes et les noyaux de nos cellules.

6 Segmentation couleur

Avant de préciser notre stratégie de segmentation, rappelons quelques notions sur la couleur. La couleur d'un point est donnée par un vecteur (R,V,B), mais il existe d'autres espaces de représentation tridimensionnels basés sur des variables obtenues à partir de l'espace RVB. On consultera [8,12,18,19] pour plus de détails sur les espaces couleurs et leurs propriétés. Dans notre segmentation nous utiliserons les systèmes RVB, HSL et $L^*u^*v^*$. Ils sont obtenus de la façon suivante :

$$H = a \cos \left[\frac{((R-V) + (R-B))}{2\sqrt{(R-V)^2 + (R-B)(V-B)}} \right] \text{ et si } V < B : H = 2\pi - H$$

$$S = 1 - \frac{3\min(R,V,B)}{L}$$

$$L = R + V + B$$

$$L^* = 116 \left(\frac{Y}{Y_0} \right)^{\frac{1}{3}} - 16 \quad \text{si } \left(\frac{Y}{Y_0} \right) > 0.008856$$

$$L^* = 903.3 \left(\frac{Y}{Y_0} \right) \text{ ailleurs}$$

$$u^* = 13L^* (u' - u'_0)$$

$$v^* = 13L^* (v' - v'_0)$$

$$u' = \frac{4X}{X + 15Y + 3Z} \quad v' = \frac{9X}{X + 15Y + 3Z}$$

$$u'_0 = \frac{4X_0}{X_0 + 15Y_0 + 3Z_0} \quad v'_0 = \frac{9X_0}{X_0 + 15Y_0 + 3Z_0}$$

$$\text{avec } \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2.76 & 1.7518 & 1.13 \\ 1 & 4.5907 & 0.0601 \\ 0 & 0.565 & 5.5943 \end{pmatrix} * \begin{pmatrix} R \\ V \\ B \end{pmatrix} \text{ et } \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

d'après [12,18].

Le système HSL a la particularité d'être plus intuitif et de permettre une interprétation plus proche de la vision humaine. Le système $L^*u^*v^*$ est perceptuellement uniforme c'est-à-dire que si deux couleurs A et B sont aussi proches entre elles que deux autres couleurs C et D, alors la différence perçue entre A et B sera équivalente à celle perçue entre C et D (ce qui n'est pas forcément vrai pour les autres espaces de représentation). L'utilisation de la couleur en analyse d'images est relativement récente [9,18]. En cytologie, on trouve quelques travaux tirant parti de l'information couleur, mais elle est utilisée plutôt pour caractériser une texture [4] que pour segmenter une image [3].

6.1 La segmentation cytoplasmique

Nous avons développé une méthode de segmentation morphologique basée sur l'utilisation de la ligne de partage

des eaux (L.P.E) [1] appliquée aux images couleurs. Les points clés de la segmentation par L.P.E sont :

- l'extraction de marqueurs des régions à segmenter (les régions à l'intérieur de chaque objet),
- la détermination de l'image sur laquelle on calcule la L.P.E,
- le calcul de la L.P.E.

La ligne de partage des eaux trouve les lignes de séparation de plus grande intensité dans une image.

6.1.1 Les marqueurs cytoplasmiques

Les marqueurs permettent d'éliminer certains minima régionaux qui pourraient être gênants. Notre but est de trouver une composante connexe par cytoplasme et d'imposer à la ligne de partage des eaux ces marqueurs.

On conviendra que :

- f = l'image couleur RVB,
- g = l'image couleur HSL,
- $E(g_L)$ = lissage exponentiel de l'image de la luminance L de l'espace HSL,
- $S(E(g_L))$ = seuillage de l'image de luminance lissée,
- $S(g_H)$ = seuillage de l'image de teinte,
- g^* = le gradient de l'image couleur f ,
- M_c = les marqueurs cytoplasmiques.

Pour trouver les marqueurs, il nous a fallu étudier toutes les composantes dans plusieurs espaces couleurs (RVB, XYZ, HSL, $L^*u^*v^*$, $L^*a^*b^*$) afin de déterminer si une de leurs composantes permet d'isoler les marqueurs cytoplasmiques de façon satisfaisante, robuste et reproductible. Il est ressorti de cette étude que seule la luminance le permettait (la luminance est la sensation selon laquelle une zone émet plus ou moins de lumière, elle est traduite par un adjectif comme claire ou foncée). Dans le cas de nos images, nous avons un fond plus clair que le cytoplasme et la transition est bien marquée au niveau intensité lumineuse. Nous avons donc basé notre détermination des marqueurs sur la luminance de l'espace HSL (composante L) et afin d'avoir des marqueurs plus précis et d'éviter les variations résiduelles dues au bruit, nous appliquons un lissage exponentiel ($E(g_L)$). Une technique de seuillage automatique est ensuite appliquée en prenant le maximum de la variance et de la vallée de l'histogramme (on obtient $S(E(g_L))$). La vallée est la première concavité que l'on trouve avant le maximum. L'histogramme de $E(g_L)$ est bi-modal et sa forme permet la reproductibilité des résultats. Cette manière de trouver les marqueurs basée sur la luminance isole de façon robuste les cytoplasmes mais parmi ces marqueurs on retrouve les hématies présentes sur les images à fond hémorragique (cf § 5). Pour y pallier nous avons utilisé la composante H (Hue : teinte) de l'espace couleur HSL. la teinte est la sensation selon laquelle une zone apparaît similaire à une des couleurs perçues : rouge, jaune-vert, bleu ou une combinaison de deux d'entre elles, elle est donnée par un angle qui permet de savoir dans quelle zone de couleur un point de couleur se situe. La figure 5 nous en donne une représentation :

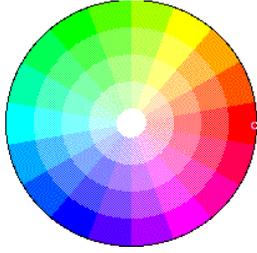


Figure 5 : la teinte dans l'espace couleur HSL.

Une seule sélection des points entre 0 et $\pi/3$ et entre $3\pi/2$ et 2π suffit pour sélectionner les objets rouges dans une image. Les valeurs permettant de sélectionner ces angles ont été déterminées en mesurant la teinte de plusieurs globules rouges afin de connaître l'intervalle de variation de la teinte pour ces derniers. L'image $S(g_H)$ obtenue subit ensuite une ouverture afin d'éliminer les irrégularités de contour : image $O^B(S(g_H))$ ou B est un élément structurant carré. A ce stade il ne nous reste plus qu'à calculer l'intersection de l'image des marqueurs des cytoplasmes avec ceux des cellules qui ne sont pas rouges : l'image $Mc=S(E(g_L)) \cap O^B(S(g_H))$ nous donne les marqueurs cytoplasmiques des cellules sans les cellules de globules rouges. Les marqueurs cytoplasmiques et le fond sont érodés plusieurs fois afin d'augmenter leur zone d'influence et permettre un meilleur partitionnement des régions.

6.1.2 Le gradient et la ligne de partage des eaux pour le cytoplasme

Une fois nos marqueurs trouvés, nous pouvons calculer la ligne de partage de eaux sur une fonction de potentiel. Nous avons choisi le gradient couleur dans l'espace RVB. La propriété du gradient est qu'il est élevé pour des points au voisinage desquels l'image présente une forte discontinuité et il est faible pour les points au voisinage desquels l'image varie peu. Cela présente un avantage certain pour nous puisque nos images ont des transitions de couleurs très franches (donc avec des discontinuités) au niveau du fond-cytoplasme. Le fond est homogène et blanc tandis que le cytoplasme est vert et hétérogène. Le gradient calculé sur des critères de couleur sera donc élevé au niveau de la transition fond-cytoplasme et favorisera la ligne de partage de eaux. Pour calculer le gradient nous utilisons un filtre 3x3 sur chacune des trois composantes de l'image couleur dans l'espace RVB :

$$\|\nabla g(x, y)\| = \sqrt{\left(\|\nabla g_R(x)\| + \|\nabla g_V(x)\| + \|\nabla g_B(x)\|\right)^2 + \left(\|\nabla g_R(y)\| + \|\nabla g_V(y)\| + \|\nabla g_B(y)\|\right)^2}$$

et l'on prend le maximum du module du gradient suivant sa direction. Les filtres utilisés suivant x et y sont :

$$FX = \frac{1}{12} \begin{pmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ -1 & -4 & -1 \end{pmatrix} \quad FY = \frac{1}{12} \begin{pmatrix} 1 & 0 & -1 \\ 4 & 0 & -4 \\ 1 & 0 & -1 \end{pmatrix}$$

Après étiquetage des marqueurs, ceux-ci constituent des régions qui vont croître dans leur zone d'influence jusqu'à un maxima du gradient : c'est la ligne de partage de eaux. Les résultats que nous avons obtenus sont très bons ; dans tous les cas les cytoplasmes sont isolés et leur contour bien déterminé. Toutes les techniques utilisées sont résumées dans la Figure 6.

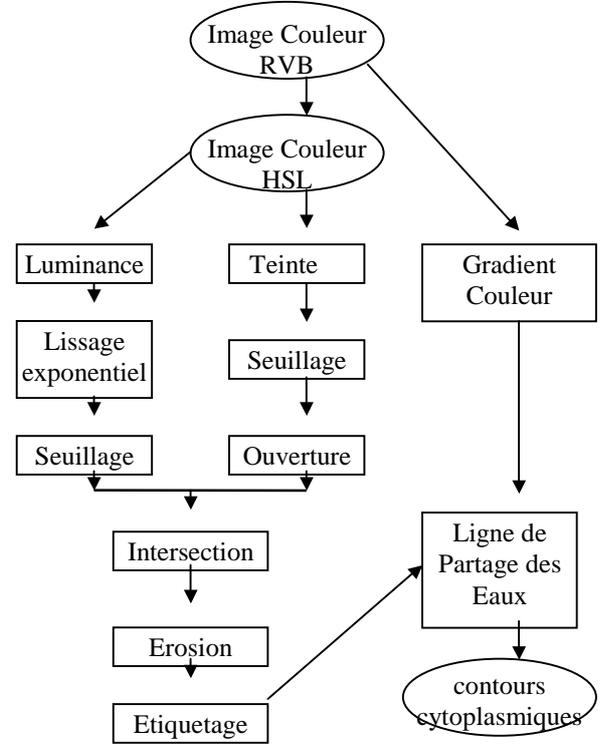


Fig. 6. Plan de la segmentation couleur des cytoplasmes.

6.2 La segmentation nucléaire

La segmentation des noyaux est également basée sur une méthode morphologique couleur basée sur la ligne de partage de eaux contrainte par marqueurs.

6.2.1 Les marqueurs nucléaires

Pour trouver ces marqueurs (soit une composante connexe par noyau) il nous a fallu étudier toutes les composantes des différents espaces couleurs et trouver celle qui peut permettre de distinguer nos noyaux du cytoplasme avec le reste de l'image. Cependant, à cause des irrégularités de densité chromatinienne dans les noyaux, certains noyaux sont très clairs et d'autres très foncés, ce qui entraîne des difficultés dans l'établissement de marqueurs isolant à la fois les noyaux clairs et foncés.

Quelques notations tout d'abord :

- f = l'image couleur RVB,
- f^* = l'image couleur RVB lissée,
- g = l'image couleur HSL obtenue par f^* ,
- g' = l'image couleur $L^*u^*v^*$ obtenue par f^* ,
- h = l'image de niveaux de gris $(g'_{v^*} + (f^*_B - f^*_V))/2$,
- $S(h)$ = seuillage de l'image h ,
- $S(g_H)$ = seuillage de l'image de teinte,
- i = le gradient de l'image $(f^*_B - f^*_V)$,

• Mn = les marqueurs nucléaires.

Nous avons alors cherché tout d'abord à minimiser les irrégularités chromatiniennes dues au bruit par un lissage exponentiel. Ensuite, il nous a fallu trouver une image permettant d'isoler de façon reproductible et robuste nos marqueurs nucléaires. Pour cela nous utilisons deux composantes couleurs : la composante v^* de l'espace $L^*u^*v^*$ qui a pour particularité de faire ressortir les noyaux clairs et la composante $(f_B^* - f_V^*)$ qui a la propriété de faire ressortir les noyaux foncés et clairs mais de façon trop agglomérée. Nous avons choisi la différence entre la composante Bleue et la Verte car nos noyaux sont bleus et le cytoplasme vert. Cette soustraction fait davantage ressortir les noyaux par rapport au cytoplasme et nivelle les différences de teinte entre les noyaux. Pour trouver ces marqueurs nous utilisons l'image $(g_{v^*} + (f_B^* - f_V^*)) / 2$, cela permet de faire ressortir à la fois les noyaux clairs et foncés privilégiant ainsi l'établissement de bons marqueurs. Pour les trouver, nous procédons par un seuillage : nous calculons la variance et la vallée de l'histogramme de h et nous prenons le seuil qui minimise la surface cytoplasmique. La composante Verte d'un point de l'image couleur est plus élevée dans le cytoplasme que dans le noyau (étant donné que le cytoplasme est vert). Nous nous servons de cette propriété pour trouver le seuil qui donne les meilleurs marqueurs. Ceci donne, avec $S1$ le seuil obtenu pour la variance, $S2$ le seuil obtenu pour la vallée et p un point d'une image :

$$\left\{ \begin{array}{l} T1 = \left(\sum f_V^*(p) \right) / p \in h \text{ et } p < S1; \\ T2 = \left(\sum f_V^*(p) \right) / p \in h \text{ et } p < S2; \\ S = S1 \text{ si } T1 < T2 \text{ et } S = S2 \text{ autrement} \end{array} \right.$$

Grâce à cette méthode, les marqueurs isolent aussi bien les cellules claires que les cellules foncées. Mais les marqueurs présentent le même inconvénient que lors de la segmentation cytoplasmique : les globules rouges en font partie. On procède donc à leur élimination par la méthode exposée au § 5.2.2 et l'on obtient nos marqueurs Mn par l'intersection entre $S(h)$ et $O^B(S(g_H))$. Ces marqueurs et le fond sont érodés plusieurs fois pour augmenter leur zone d'influence.

6.2.2 La Ligne de Partage de Eaux sur les noyaux

Le gradient est calculé à partir de l'image $i = (f_B^* - f_V^*)$: comme elle nivelle les différences entre les noyaux et accentue la différence noyau-cytoplasme, le gradient est élevé à l'endroit de la transition bleu-vert (noyau-cytoplasme). Après étiquetage des marqueurs Mn , on calcule la Ligne de Partage des Eaux. La stratégie entière est résumée dans la figure 9. Les résultats obtenus sont visuellement bons : les noyaux des cellules isolées bien définis, les noyaux accolés sont détectés (Figure 7) comme tels et les amas forment des zones plus complexes mais formant bien un seul objet (Figure 8).

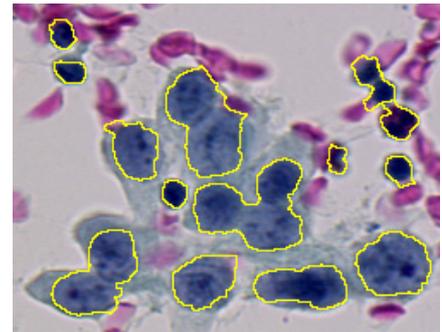


Figure 7 : des noyaux accolés.

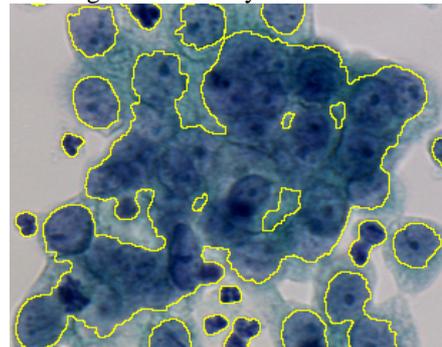


Figure 8 : des noyaux en amas.

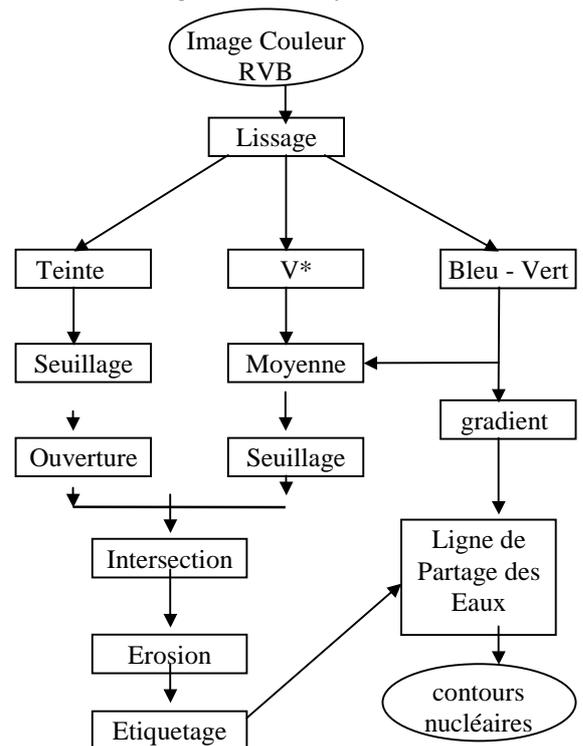


Figure 9 : Plan de la segmentation couleur des noyaux.

6.3 Séparation des noyaux accolés

Notre méthode de segmentation des noyaux permet de déterminer leurs bordures nucléaires, mais les noyaux accolés ne sont pas séparés (Figure 7). Il faut donc une étape de plus pour les dissocier. Nous avons utilisé encore une fois la ligne de partage de eaux contrainte par marqueurs. Cependant, cette fois, la fonction potentielle est l'inverse de la fonction distance aux contours nucléaires. Nous avons utilisé une distance de chanfrein 3x3. Les marqueurs sont les érodés ultimes de chacun des noyaux obtenus à la précédente étape (§5.3). Ces érodés sont calculés par une reconstruction par érosion de hauteur 2. Les marqueurs obtenus permettent alors de séparer plusieurs noyaux grâce à l'inverse de la fonction distance. En effet, prenons deux noyaux se touchant. La fonction distance est élevée pour les points au centre de chaque noyaux (Figure 10). En prenant l'inverse on peut alors séparer le noyau en deux parties qui correspondent aux deux noyaux accolés c'est-à-dire les minima régionaux donnés par les érodés ultimes. Un exemple de séparation des noyaux est donné par la Figure 11 : les lignes rouges donnent la séparation trouvée entre deux noyaux qui étaient confondus. Cette méthode s'avère fonctionner pour les cas qui nous intéressent : les cellules accolées présentant une concavité dans leur bordure nucléaire. Dans les autres cas, les noyaux ne sont pas séparables du point de vue cytologique (chevauchement des noyaux) et ils ne doivent pas être séparés. La stratégie de séparation est résumée par la Figure 12.

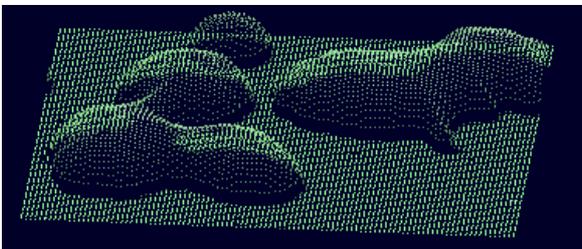


Figure 10 : La fonction distance sur des noyaux accolés en vue 3D.

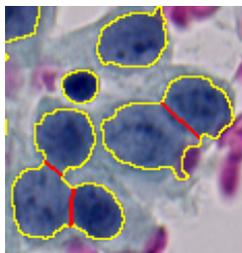


Figure 11 : Les noyaux après séparation.

7 Discussion

Les résultats que nous avons obtenus sont très encourageants. Les figures 13b et 13c présentent les cytoplasmes et les noyaux extraits de l'image 13a (les bordures cytoplasmiques et nucléaires sont en jaune). L'image 13d présente la séparation des noyaux accolés, les

lignes rouges sont les frontières entre les noyaux séparés. Nous avons testé notre segmentation sur 80 images dont 50 prises au hasard sur une lame et 30 sélectionnées parce qu'elles présentaient une difficulté cytologique (amas, globules rouges, types cellulaires très variés, papilles). Pour toutes nos images, les contours cytoplasmiques et nucléaires sont bien déterminés. L'évaluation des résultats a été faite de manière visuelle par les pathologistes et les experts en analyse d'images. Cette évaluation consistait à vérifier la bonne détermination des bordures cytoplasmiques et nucléaires ainsi que la segmentation effective de toutes les cellules de l'image. Nous avons donc mis au point la première étape de notre système ARCTIC qui consiste à isoler les cellules sur une image. Cette étape basée sur l'utilisation des informations couleurs est robuste et reproductible. Sa mise au point a demandé la collaboration de connaissances pluridisciplinaires. L'utilisation de la couleur dans une segmentation couplée à des outils morphologiques puissants offre au pathologiste une technique de reconnaissance plus sûre des cellules sur une image de cytologie des séreuses.

Remerciements

Ce travail est financé par *La Ligue Contre le Cancer (Comité de la manche)* dans le cadre d'une thèse de doctorat et est réalisé au sein du *Pôle Traitement et Analyse d'Images de Basse-Normandie*. Il est mis au point plus précisément au *Laboratoire Universitaire des Sciences Appliquées de Cherbourg* et au *Service d'anatomie et de Cytologie Pathologiques de l'Hôpital Pasteur de Cherbourg*.

Annexe

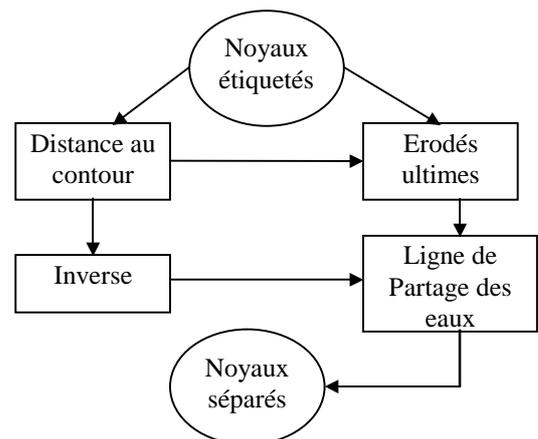


Figure 12 : La stratégie de séparation des noyaux.

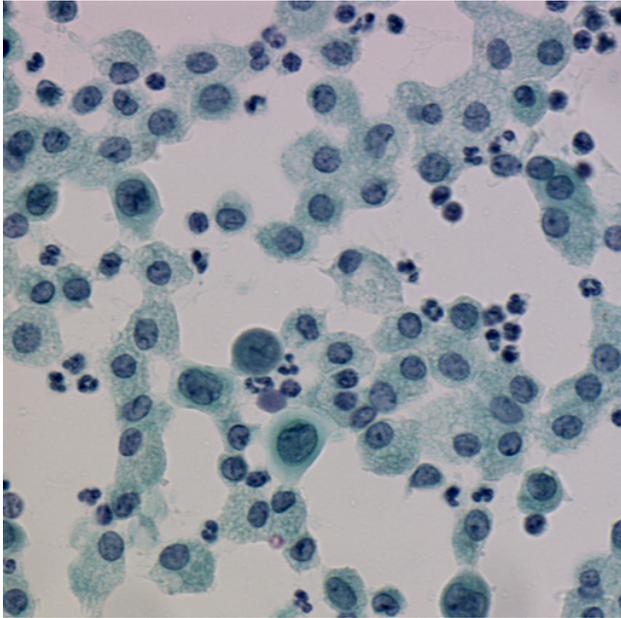


Figure 13.a : Image Couleur de cellules provenant de liquides d'épanchement des séreuses.

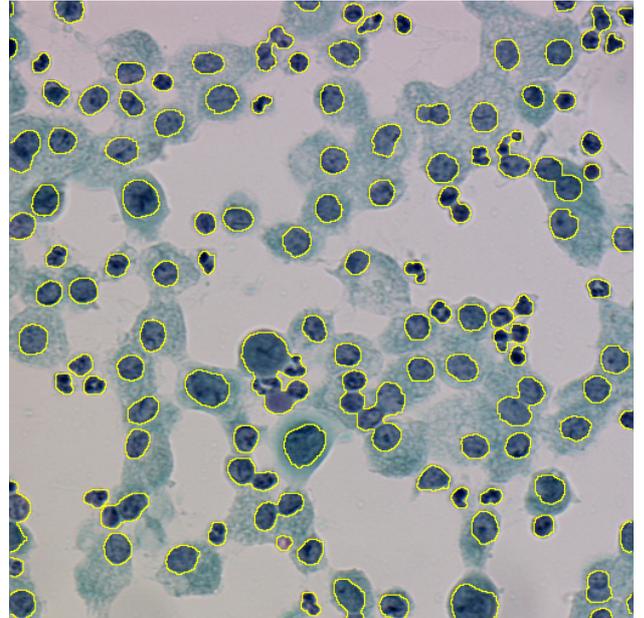


Figure 13.c : Image des contours nucléaires avant séparation.

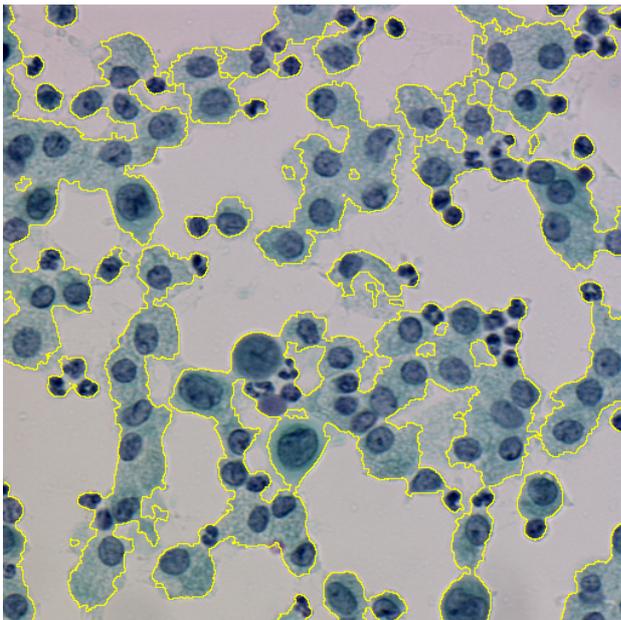


Figure 13.b Image des contours cytoplasmiques.

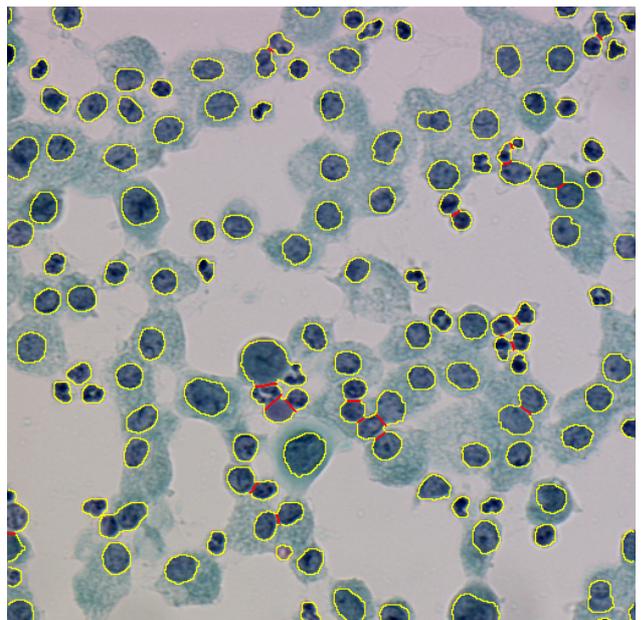


Figure 13.d : Image des contours nucléaires après séparation.

Figures 13 : les images des cellules des séreuses à chacun des différents stades de la segmentation.

Références

[1] S. Beucher, *Segmentation d'images et morphologie mathématique*, Ph.D thesis, Ecole Nat. Sup. des Mines de Paris, 1990.

[2] R. Clouard, A. Elmoataz, F. Angot, O. Lezoray, *Pandore : une bibliothèque et un environnement de*

programmation d'opérateurs de traitement d'images, Rapport interne du GREYC, Caen, France, Mars 1997.

[3] C. Garbay, Image structure representation and processing : a discussion of some methods in cytology, *IEEE Trans. On PAMI*, Vol. 8, n°2, pp. 140, Mars 86.

- [4] H. Harms, U. Günzer, H.M Aus, Combined local color and texture analysis of stained cells, *CVGIP*, Vol. 33, pp. 364-376, 1986.
- [5] IAC, International Academy of cytology, Specifications for automated cytodiagnostic systems proposed by the I.A.C, *Acta Cytologica*, Vol.28, pp. 352, 1984.
- [6] L.G Koss, The Papanicolaou test for cervical cancer detection : a triumph and a tragedy, *J. Am. Med. Assoc*, Vol. 261, pp. 737-743,1989.
- [7] E.A Knesel Jr., Roche Image Analysis System, *Acta Cytologica*, Vol. 40, n°1, pp. 60-66, 1996.
- [8] M. Kunt, *Traitement de l'information numérique des images*, Presses polytechniques et Universitaires romanes, 1991.
- [9] Q.T. Luong, *La couleur en vision par ordinateur : 1- Une revue* , Rapport de recherche INRIA n°1251, 1990.
- [10] G.N Papanicolaou, A new procedure for staining vaginal smears, *Science* , Vol. 95, pp.432, 1942.
- [11] S.F Patten Jr., S.S.J. Lee, A.C. Nelson, NeoPath AutoPap 300 Automatic Pap Screener System, *Acta Cytologica*, Vol. 40, n°1, pp.45-52, 1996.
- [12] H. Pauli, Proposed extension to the CIE recommendations on uniform color spaces, color difference equations and metric color terms , *J. Opt. Soc. Am.*, Vol. 66, No 8, pp. 866-867,1975.
- [13] M. Revenu, A. Elmoataz, C. Porquet, H. Cardot, An automated system for the classification of cellular categories in cytological images, *Intelligent Robot and Computer Vision XII, Algorithms and Techniques*, SPIE, Boston, USA, pp. 967-970, Sept 93.
- [14] *Revision of Laboratory requisitions for medicine, medicaid and clinical laboratories improvement act of 1967 CLIA'67 programs*, Federal Register, Vol. 5, pp. 9600, 1967.
- [15] D.L Rosenthal , D. Acosta, R.K Peters, Computed-Assited Rescreening of Clinically Important False Negative Cervical Smears Using the PAPNET Testing System, *Acta Cytologica*, Vol. 40, n°1, pp. 120-126, 1996.
- [16] N. Tanaka, H. Ikeda, T. Ueno, A. Mukawa, S. Watanabe, K. Okamoto, S. Hosoi, S. Tsunekawa, Fundamental study of cyto-screening for uterine cancer : new system for automated apparatus (CYBEST) utilizing pattern recognition method, *Acta Cytologica*, Vol. 21, pp. 85, 1977.
- [17] W.E Tolles, The Cytoanalyser : an example of physics in medical research, *Trans. N.Y.Acad.Sc*, Vol. 17, pp. 250, 1955.
- [18] A. Tremeau, *Contribution des modèles de la perception visuelle à l'analyse d'images couleurs* , Thèse de l'université de Saint-Etienne, 1993.
- [19] G.Wyszecki, W.S Styles, *Color Science : Concepts and Methods, Quantitative Data and Formulae*, Second Edition , John Wiley & Sons, New-York, 1982.